

Phoneme discrimination using KS -algebra II.

Ondrej Šuch*, Lenka Mackovičová†

February 26, 2013

Abstract

KS -algebra consists of expressions constructed with four kinds of operations, the minimum, maximum, difference and additively homogeneous generalized means. Five families of Z -classifiers are investigated on binary classification tasks between English phonemes. It is shown that the classifiers are able to reflect well known formant characteristics of vowels, while having very small Kolmogoroff's complexity.

1 Introduction

In our previous paper in the series we have proposed a new KS -algebra for constructing binary phoneme classifiers based on spectral content. The algebra consists of expressions constructed from a vector of spectral values $\mathbf{s} = (s_1, \dots, s_n)$, and the zero value by means of the following operators

*O.Šuch is with Slovak Academy of Sciences, Banská Bystrica, Slovakia, ondrejs@savbb.sk

†L. Mackovičová is with University of Matej Bel, Banská Bystrica, Slovakia, lenka.mackovicova@umb.sk

‡Work on this paper was partially supported by research grant VEGA 2/0112/11; Computations were done on computers purchased in project ITMS code: 26210120002

- the minimum $\min(x_1, \dots, x_n)$,
- the maximum $\max(x_1, \dots, x_n)$,
- the difference $x_1 - x_2$,
- the additively homogeneous means A_α ,

where

$$A_\alpha(x_1, \dots, x_n) = \ln\left(M_\alpha(\exp(x_1), \dots, \exp(x_n))\right),$$

and M_α is the generalized mean

$$M_\alpha(x_1, \dots, x_n) = \left(\frac{x_1^\alpha + \dots + x_n^\alpha}{n}\right)^{1/\alpha}.$$

In this article we shall present results of search for optimal Z -classifier in a large, albeit special family of elements of KS -algebra.

2 Optimization setup

For dataset we shall use spectral data presented in [1] and used for demonstration in [2] and [3]. The data is derived from TIMIT database, often used in speech recognition tasks. It consists of 5 English phonemes, three vowels **aa**, **ao**, **iy** and two consonants **dc1**, **sh**, each pronounced by a male speaker from various geographical regions. The sound was sampled at 16kHz, and spectral data was prepared using 512-sample window, resulting in 256

spectral vector for each sample. The data is divided into train and test categories, with approximately equal proportions in each.

Let us recall that a general Z -classifier for phonemes ϕ_1, ϕ_2 corresponds to an element f of KS -algebra. Suppose the classifier is presented with spectral data \mathbf{s} and prior knowledge that the data corresponds to either ϕ_1 or ϕ_2 . It decides that phoneme is ϕ_1 if $f(\mathbf{s}) < 0$ and decides that the phoneme is ϕ_2 if $f(\mathbf{s}) > 0$.

For an optimization criterion we chose the number of successful classifications $c(f)$ on the training data set. Since the data set is rather small, ties may occur. In the case of ties, we choose the classifier that maximizes the expression

$$\rho(f) := \min\left(\frac{\mu_1^2}{\sigma_1^2}, \frac{\mu_2^2}{\sigma_2^2}\right),$$

where μ_i, σ_i are sample means and standard deviations for values of f on the set of training samples of phoneme ϕ_i . The expression plays role analogous to that of Fischer's linear discriminant.

Since there is no obvious shortcut to finding an optimum, we resort to evaluating classification performance in turn for every classifier in a given family.

3 Families of classifiers

By a *spectral range* we mean a sequence $R_{i,j} = (s_i, s_{i+1}, \dots, s_j)$ of consecutive spectral amplitudes (ordered by increasing frequency). Since brain structures devoted to speech recognition are tonotopically organized [4], we

propose to use for discrimination functions defined on spectral ranges. Each discrimination function takes the form

$$f = f_1(R_{i,j}) - f_2(R_{k,l}),$$

where f_1, f_2 are symmetric, additively homogeneous functions of KS -algebra. The difference of values of f_1 and f_2 is then intensity invariant. We distinguish five different classes of such functions

1. the mean of values in the spectral range
2. the mean of m largest values in the spectral range
3. A_1 average of m largest values in the spectral range
4. A_2 average of m largest values in the spectral range
5. a quantile of the spectral range (the m -th largest value)

Obviously, family 1 is a subset of family 2. Families 2–5 can be seen as special cases of a family obtained by taking average A_α of m largest values ($\alpha = 0, 1, 2, -\infty$ respectively). Families 1, 2 and 5 are special cases of so called OWA-operators [5]. A general n -ary OWA operator F with weight vector $\mathbf{w} = (w_1, \dots, w_n)$ is defined by expression

$$F(a_1, \dots, a_n) = w_1 b_1 + \dots + w_n b_n,$$

where b_i is the i -th largest element of the set $\{a_1, \dots, a_n\}$.

If one limits oneself to searching over pairs of functions defined on ranges of width up to w , the complexity of selecting the best one is $\sim kNw^6$ in

families 2–5, and $\sim k'Nw^4$ in family 1, where N is the number of samples in the training set. Despite using optimized software, the former growth is still quite large, and we opt to search only through a fixed number of values of m in families 2–5 that includes $m = 1$, $m = w$, and m close to $w/4$, $w/2$ and $3w/4$ respectively.

4 Trainability

Trainability of classifiers is their ability to capture class distributions over training data. Significant failure to classify training data is an indication that the classifier is not flexible enough. Training errors for each class of classifiers is shown in Table 1.

	1	2	3	4	5
aa-ao	232	213	223	224	220
aa-dcl	1	1	0	1	1
aa-iy	0	0	0	0	0
aa-sh	1	0	0	0	0
ao-dcl	1	1	1	1	1
ao-iy	0	0	0	0	0
ao-sh	1	0	0	0	0
dcl-iy	69	47	71	73	73
dcl-sh	1	0	0	0	0
iy-sh	8	0	0	0	1

Table 1: Total number of errors on training data of the best classifiers in the family given in the column for a pair of phonemes given by the row

From the table we can see that family 1 of classifiers is least trainable, which can be expected, since it is subsumed by family 2. On the other hand, family 2 is slightly more trainable than others.

5 Performance on test data

The crucial characteristic of any classifier is its performance on the test data.

Table 2 summarized results of best classifiers within a given family (priority is given to $c(f)$ and $\rho(f)$ is used in case of ties). We can again see that

	1	2	3	4	5	Family 2 pctg.
aa-ao	95	91	96	98	94	79.95%
aa-dcl	0	0	0	0	0	100%
aa-iy	0	0	0	0	0	100 %
aa-sh	2	1	0	0	0	99.75%
ao-dcl	0	0	1	0	0	100%
ao-iy	0	0	1	2	1	100 %
ao-sh	1	0	0	0	0	100 %
dcl-iy	30	16	24	25	34	96.84%
dcl-sh	1	1	0	0	0	99.76%
iy-sh	5	1	2	2	1	99.81%

Table 2: Total number of errors on test data of the best classifiers in the family given in the column for a pair of phonemes given by the row

family 2 of classifiers provides the best testing performance. Intriguing is poor performance of family 5 on discrimination of dcl versus iy. In fact, there is a simple classifier in family 5 with only 24 errors on test data. Putting priority on the correct train count $c(f)$ rather than on $\rho(f)$ resulted in reporting performance of poorer classifier (dcl.iy.2.disc in R code listing below) rather than the better one (dcl.iy.1.disc).

```
dcl.iy.1.left.1 = function(x) max(x[2:6])
dcl.iy.1.right.1 = function(x) max(x[10:13])
dcl.iy.1 = function(x) dcl.iy.1.left.1(x) - dcl.iy.1.right.1(x)
dcl.iy.1.disc = function(x) if (dcl.iy.1(x) < 0) "iy" else "dcl"
```

```

dcl.iy.2.left.1 = function(x) max(x[1:4])
dcl.iy.2.right.1 = function(x) Q(x,8,14,5)
dcl.iy.2 = function(x) dcl.iy.2.left.1(x) - dcl.iy.2.right.1(x)
dcl.iy.2.disc = function(x) if (dcl.iy.2(x) < 0) "iy" else "dcl"

```

```

Q = function(x,a,b,p) { v = sort(x[a:b]); return(v[p]) }

```

6 Visualization

The families of discriminators we have examined in this article can be readily visualized.

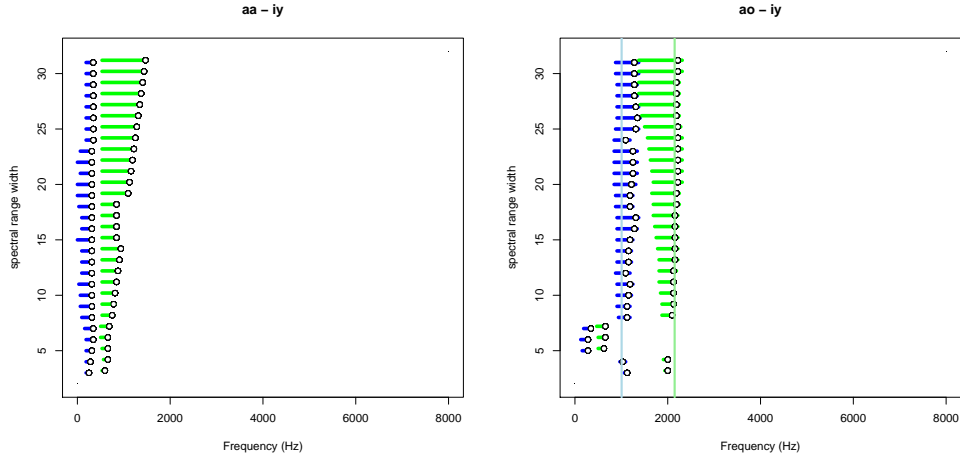


Figure 1: Supports and position of parameter m (white circles) of optimal classifiers of a given width in family 5. In the right picture locations of average frequencies for formats F2 for ao (light blue) and iy (light green) are indicated by vertical lines. The average values were taken from [6].

In Figure 1 we can see how support of f_1 and f_2 changes if we allow

increased size of its support. In the right picture we can see that support of components of f matches well with formants.

7 Conclusion

We have conducted a search for structure of optimal Z -classifiers in 5 families of functions in KS -algebra. Among families we considered, slightly better results on both training and test data were obtained in family 2. We have demonstrated that classifiers found by our procedure reflect well known formant concept. Advantages of these classifiers include clear interpretation, visualizations and lack of any continuously varied parameters resulting in low Kolmogoroff's complexity.

Further research should investigate more general classes of KS -algebra based classifiers, namely B and A -classifiers, adjustments for psychoacoustic phenomena, and develop means to compose single feature classifiers.

References

- [1] English phonemes.
URL <http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/phoneme.data>
- [2] T. Hastie, R. Tibshirani, J. Friedman, Elements of statistical learning, Springer.
- [3] T. Hastie, A. Buja, R. Tibshirani, Penalized discriminant analysis 23 (1) 73–102.

[4] Tonotopy.

URL <http://en.wikipedia.org/wiki/Tonotopy>

[5] R. R. Yager, An ordered weighted averaging aggregation operators in multicriteria decisionmaking, IEEE Transactions on systems, man and cybernetics 18 (1) (1988) 183–190.

[6] H. Sharifzadeh, I. McLoughlin, M.J.Russell, A comprehensive vowel space for whispered speech, Journal of voice 26 (2) (2012) e49–e56.